

시맨틱 웹의 응용 사례 연구

한국과학기술원 조성정 · 김진형*

1. 서론

팀 버너스-리(Tim Berners-Lee)가 1990년에 제안한 웹은 지난 10여 년간 규모와 사용 환경에서 급속하게 성장하였다. 웹 공간상에 참여하는 웹사이트와 사용자의 수는 기하급수적으로 팽창하였다. 또한 HTML을 기반으로 작성한 페이지들이 링크로 연결된 정적 환경으로부터 필요에 따라 정보가 생성되며 사람뿐만 아니라 소프트웨어 에이전트도 참여하는 CGI, JAVA와 웹서비스 기반의 동적 환경으로 진화하였다[10].

사용자가 문서를 읽기 쉽도록 정보를 보여주는 데 중점을 둔 기존의 웹 기술로는 양적, 질적으로 변화한 웹 공간을 지원하는데 어려움이 존재한다[1,4,5,10]. 대부분의 정보는 화면상에 문서의 모양을 정의하는 태그로 구성된 HTML로 작성되며, 자연어로 기술된다. 사람이 문서를 이해 하는데는 문제가 없으나, 기계가 문서의 의미를 이해하고 처리하는데 한계가 있다. 따라서 소프트웨어 에이전트와 관계된 정보 교류, 즉 에이전트와 사람간, 에이전트와 에이전트간 정보와 지식을 교류하는데 한계가 있다. 또한 현재의 정보 검색은 단순하게 키워드를 통하여 이루어지므로, 동일한 단어지만 다른 의미를 갖는 문서가 검색되고, 동일한 의미지만 다른 단어로 표현된 문서는 검색되지 않는 문제가 종종 발생한다.

이에 대한 해결책으로 팀 버너스-리는 1998년에 기계가 정보를 이해하고 처리하도록 기존 웹을 확장하는 것을 목표로 하는 시맨틱 웹을 제안하였다[1,4,5,10]. 시맨틱 웹에서는 자원(문서, 정보, 개념 등)을 유일한 이름(URI)로 지칭하고, 명시적으로 온톨로지를 통하여 표준화된 방식으로 지식을 공유하

고 교환할 것을 제안하고 있다. 기계가 정보의 내용을 이해함으로써, 정보를 보다 정밀하게 선별하고, 지능적인 분류 및 검색이 가능하다. 또한 기계가 온톨로지에 기반한 추론을 통하여 암묵적으로 내재된 지식도 처리할 수 있다.

시맨틱 웹은 정보와 지식의 표현, 교환, 축적 및 검색이 중요한 의미를 갖는 분야이거나, 웹 서비스 같이 소프트웨어의 서비스 내용 검색 및 파라미터 표준화가 중요한 분야와 에이전트간 지능적인 정보 교환이 중요한 분야에 적합하다. 본 논문에서는 성공적인 응용 사례로서 기존 지식관리 시스템에 온톨로지의 추출, 생성, 관리 및 검색도구를 추가하는 On-To-Knowledge 과제[12], 전자상거래를 위한 상품의 분류체계인 UNSPSC[14], 음악 분야에서 가수, 곡 제목, 앨범 제목, 트랙번호 등의 메타데이터를 교환하고 분류하는 MusicBrainz[11], IT 관련 세미나 정보를 관심있는 사용자에게 선별하여 제시하는 ITTA LKS[2], 국제 공동 연구기관들의 컴퓨팅 자원을 하나로 묶어 장시간에 걸쳐 실험을 수행하는 그리드 컴퓨팅[9] 등의 사례를 살펴보고자 한다.

2. 시맨틱 웹의 구조[4,5,10]

시맨틱 웹은 그림 1과 같이 계층적인 구성요소로 이루어진다. 가장 기본이 되는 층은 자원을 명시적으로 지칭하는 URI와 UNICODE로 구성되어 있다. 다음 층은 임의의 개념을 모듈방식으로 정의할 수 있는 XML(Extensible Markup Language)과 명칭공간(Namespace)이며, 그 다음으로 자원을 기술하기 위한 RDF(Resource description framework)가 위치한다. 상위에는 온톨로지를 지원하고, 규칙, 논리, 증명을 위한 기술 요소의 층들이 위치한다.

XML은 각자 웹 페이지의 일부분을 사용자가 정

* 중신회원

의한 태그로 표시하는 것이 가능하며, 사용자가 해당 문서에 대하여 임의의 구조를 추가할 수 있다. 반면 이 구조의 보편적인 의미는 정의하지 않는다.

RDF에서는 의미를 주어, 동사, 목적어에 해당하는 세 가지 요소로 표현한다[15]. 주어와 목적어에 해당하는 요소는 개념과 객체를 의미하며, 동사에 해당하는 요소는 이들의 관계를 정의한다. 예를 들어 “(영희(는), 철수(의), 동생이다)”로 표현하는 문장은 ‘영희’가 ‘철수’에 대하여 ‘동생’의 관계를 갖음을 의미한다. 각 요소들은 URI로 표현하여 유일하게 지칭함으로써, 서로 다른 개념을 동일하게 지칭할 혼동을 없앤다.

RDF+Schema 레이어에서는 클래스, 하위 클래스, 존재물(entity)간 관계가 추가되며, 속성이 속할 수 있는 클래스(domain)와 속성이 가질 수 있는 값의 범위(range)가 추가되어, 개념들간의 관계를 제한할 수 있다[16].

온톨로지는 특정 도메인의 지식을 명시적으로 표현하기 위한 개념들, 개념 사이의 관계, 개념의 속성 및 특성, 속성 및 특성에 부여된 제약조건 및 객체들로 표현된다. 이 온톨로지를 통하여 특정 도메인의 단어를 공통으로 정의하고, 지식을 공유할 수 있다. 예를 들어, 주소 및 연락처에 대한 공통의 온톨로지를 사용할 경우, 상이한 데이터베이스의 ‘이동전화 번호’ 필드와 ‘휴대폰 번호’ 필드가 동일한 개념임을 알 수 있다.

규칙(Rules)과 논리(Logic) 계층은 이러한 온톨로지에 기반하여 질의와 추론을 가능하게 한다. 상위 클래스의 속성은 하위 클래스에 상속된다는 규칙과, 그 밖의 다른 규칙들을 적용하여 암묵적으로 내재된 새로운 관계들을 생성할 수 있다. 또한 추론의 검증을 요구하는 것도 가능하다. 예를 들어 예기치 못한 사실에 대한 주장이 입력될 경우, 어떤 추론 단계를 거쳐 결론에 도달하였는지 확인을 요구할 수 있다.

3. 시맨틱 웹의 응용사례

3.1 On-To-Knowledge 과제: 온톨로지를 통한 지식 관리 도구의 개발[8,12]

On-to-knowledge(OTK) 과제는 EU의 여러 연구 기관(네덜란드: Univ. of Amsterdam, 독일: AIFB, 영국: BT lab., 노르웨이: CognIT, 스웨덴: Enersearch, 불가리아: OntoText Lab.)이 콘소시엄을 구

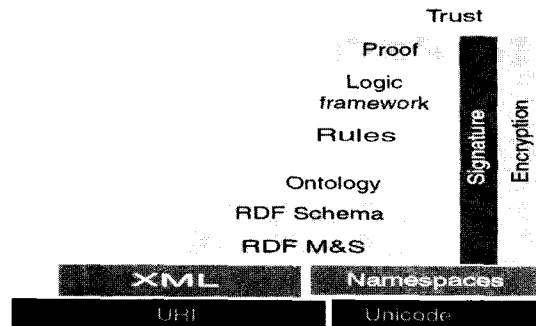


그림 1 팀 버너스-리가 제안한 시맨틱 웹의 구조[5]

성하여 온톨로지를 이용한 내용기반 지식 관리 도구를 개발하는 과제이다.

치열한 경쟁의 시대에서 기업의 생존에는 기업에서 생성되는 대용량의 지식을 효과적으로 관리하는 것이 필수적이다. 기존 지식 관리 시스템은 정보가 이질적인 형태로 다양한 물리 공간상에 산재된 경우, 체계적인 관리가 어렵고, 키워드로만 대용량의 정보를 검색하는 비효율성의 문제를 갖고 있다. OTK 과제는 기업내 정보를 시맨틱 웹 기술을 응용하여 효과적으로 관리하는 것을 목표로 하며, 크게 세 가지 분류의 기능으로 구성된다.

1. 정보의 습득 : 텍스트에서 정보의 의미를 추출하는 기술이다.
2. 정보의 유지 및 관리 : RDF, XML 및 OIL 등으로 정보의 의미를 표현한다. 여기에는 사용자가 기존 문서에 시맨틱을 추가하는 도구, 온톨로지 편집도구, 기존 문서에서 자동으로 온톨로지를 저장 및 검출하는 도구, 온톨로지 저장 DB 등이 있다
3. 접근 : 푸쉬 서비스 및 에이전트 기술로 사용자가 정보를 접근하도록 지원한다. 온톨로지에 기반한 정보 바다의 항해, 검색 및 가시화 도구가 있다.

그림 2는 OTK 과제의 전체 도구 환경을 보여주고 있다. OntoExtract와 OntoWrapper는 특정 도메인과 관련한 비정형 및 정형 문서를 추출한다. 추출한 정보는 RDF-DB(sesame)에 저장되고, OntoEdit로 편집한다[13]. 다음으로 RQL(RDF query language) 질의 엔진으로 DB의 내용을 질의하고, 결과를 RDF Ferret와 Spectacle로 사용자에게 전달하고 가시화 한다.

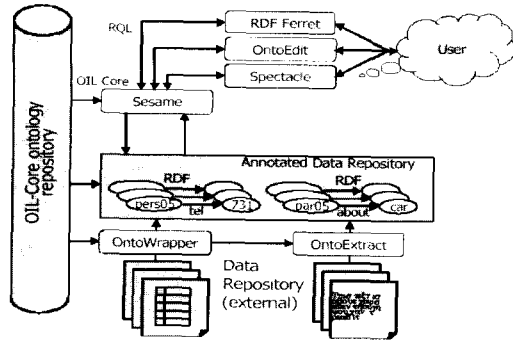


그림 2 OTK의 도구 환경[8,12]

OTK 과제를 실제 현장의 문제에 적용한 사례는 다음 두 가지를 들 수 있다.

1. British Telecom 콜 센터(call center)

회사의 콜 센터는 고객의 전화에 응답하여, 고객에 필요한 서비스를 제공하는 플랫폼이다. 콜 센터의 사원들은 고객의 문제 해결을 위하여 회사의 상품 및 서비스를 충분히 이해하고, 자주 발생하는 문제의 해결방안을 경험적 지식(practice)으로 갖추고 있다. British Telecom에서는 인트라넷 기반의 가상의 공동체(community)를 만들어서, 콜 센터의 직원들이 정보를 서로 공유하는데 온톨로지를 도입하였다. 이 공동체에서는 기존에 성공적인 경험지식을 저장하고, 새로운 경험지식을 자동으로 배포하는 기능을 수행한다. 온톨로지의 도입으로 지식을 공통의 어휘로 표현함으로써, 신입사원이 기존의 경험지식을 빠르게 학습하게 하며, 지식의 표준화된 표현이 가능하게 된다.

2. Swiss Life 생명보험회사의 계정 표준 문서 검색

Swiss Life 생명보험회사에서는 국제 계정 표준에 대한 대용량의 문서에서 관련 정보를 검색하는 업무에 온톨로지를 이용한다. OntoExtract를 이용하여 온톨로지를 추출하고, 사용자는 온톨로지를 이용하여 세밀한 질의를 수행한다. 또한 직원이 보유한 특별한 기술, 업무내용 및 학력을 온톨로지를 이용하여 작성한 후 공유한다. 이를 통하여 특정한 기술적 능력, 업무경력 및 학력을 갖춘 책임자를 검색자의 의도에 부합하게 검색하는 것이 가능하다.

3.2 UNSPSC: 전자상거래를 위한 상품 분류 체계[14]

인터넷상의 전자 상거래는 제품 판매자가 제시하

는 제안과 소비자간의 요구가 비슷하게 대응될수록 성사될 가능성이 높아진다. 또한 상거래에 직접 참여하는 판매자와 소비자 외에 가격 비교 사이트, 제품 알선 사이트 등 다양한 형태의 웹 서비스에 에이전시가 참여한다. 따라서 내용량의 다양한 상품과 서비스에 대한 정보를 소프트웨어들이 교환하고, 선별하는 것이 거래 성사에 중요한 역할을 한다.

전자카탈로그와 같이 여러 판매 사이트간 제품 정보를 통합하여 공통으로 분류하는 체계로서 UNSPSC을 들 수 있다[3,14]. 이는 전자상거래 및 비즈니스에 참여하는 모든 물품과 서비스를 계층적으로 분류하는 표준이다. 제품 분류는 간단한 온톨로지로서 표현하며, 제품에 대한 공통의 어휘와 계층적 구조로 표준화한다. 이 온톨로지를 이용하여, 여러 사이트들이 제품을 동일하게 지칭할 수 있으며, 공통의 어휘로 정보를 교환할 수 있다.

UNSPSC를 지원하는 대표적인 제품 분류 자동화 소프트웨어에는 GoldenBullet가 있다[7]. 이 소프트웨어는 현재까지 작성된 UNSPSC로 표현된 분류체계와, 분류하고자 하는 제품에 대한 간단한 설명을 입력받는다. 이 설명 내용은 정보검색, 기계학습 및 자연어 처리 기법을 이용하여, 간단한 온톨로지로 변환되며, UNSPSC에서 가장 유사도가 큰 항목으로 분류된다. 현재는 영어와 프랑스어를 지원하고 있다.

3.3 MusicBrainz: 포괄적인 음악 메타데이터베이스[11]

MusicBrainz¹⁾는 포괄적인 음악 백과사전을 목표로 하는 음악 메타데이터베이스이다. 오디오(CD)와 MP3 등의 디지털 음악 트랙을 인지하여(30초간 음악 내용을 ID로 변환), 이 곡에 대한 메타데이터(음악가, 곡 제목, 앨범 제목, 트랙 번호, 출시 날짜, 오디오 트랙 시그내처, CD disc ID 등)를 제공하는 것을 목표로 한다. 새로 구입한 CD를 오디오 재생기에 넣고 연주를 시작하는 시나리오를 예를 들어 살펴보자. 많은 경우 오디오 재생기의 정보 표시창에는 Audio CD 1, Track 1, Track 2 이런 식으로 트랙 번호가 표시된다. 그러나 오디오 재생기가 MusicBrainz에 접속할 경우에는 해당 곡의 메타데이터를 전송받아, 그 곡의 음악가, 앨범 이름, 트랙번호 등을 표시할 수 있다.

1) 2003년 2월 현재 음악가 31,876명, 앨범 53,348개, 트랙 655,229개가 등록되어 있다[11].

현재 웹 상에서는 소리바다, 냅스터, 그누텔라 등의 서비스를 이용하여 P2P로 음악파일들이 많이 공유되고 있으며, 소비자들은 또한 많은 CD를 보유하고 있다. 음악에 대한 메타데이터는 음악파일의 일부가 아니기 때문에, 각 회사 혹은 각 개인마다 각자의 방식대로 작성한다. 따라서 메타데이터를 교환하고 처리할 때 표준화의 문제가 야기된다.

MusicBrainz는 메타데이터 정보들을 간단한 온톨로지 표준화를 통해 체계적으로 관리한다. 동일한 목적으로 1996년에 Internet Compact Disc Database (CDDDB)가 작성되었다. 하지만 시맨틱 정보 표현력의 한계로 잘못된 철자와 중복된 메타데이터 내용 등의 혼란이 문제가 되고 있다. MusicBrainz에서는 음악가, 앨범 제목, 트랙 번호에 URI를 부여하고, 정보를 RDF로 기술한다. URI로 객체들이 유일하게 결정되므로, 메타데이터의 중복을 피할 수 있고, 잘못된 철자의 문제가 발생할 여지가 적다. 추가적으로, 외부의 응용 소프트웨어가 RDF로 기술된 음악 정보를 접근하고 해석하여, 자신만의 서비스를 제공할 수 있다.

3.4 ITTALKS: IT 관련 세미나의 선별 및 예약[2]

ITTALKS는 IT 관련 세미나 및 회의에 대한 정보를 종합적으로 제공하는 웹 포털을 목표로 한다. 이 시스템은 IT 관련 세미나를 도메인으로 사람과 에이전트간, 에이전트와 에이전트간 상호작용을 가능하게 하는 웹 기반 인프라스트럭처이다. 지식의 표현과 추론, 에이전트간 대화에 DAML(DARPA Agent Markup Language[6])을 사용함으로써 기존 데이터베이스의 단순한 서비스에 비하여 지능적인 서비스를 제공한다.

한 예를 들어보자. 사용자는 IT 관련 관심사와 개인 스케줄 등을 프로파일의 형태로 기술한다. 사용자 에이전트는 이 프로파일에 나타난 복잡하고 가변적인 제약조건을 만족하는 IT 관련 세미나 자료를 ITTALKS에서 선별하여 얻어온다. 세미나의 참석 가능성을 판단하기 위하여, 세미나가 진행중인 시간에 사용자의 예상 위치, 세미나 장소에서의 거리 및 해당 시간대 교통상황을 동시에 종합적으로 고려한다. 판단에 필요한 자료들은 세미나 주제 분류 에이전트, 지도 에이전트와 같은 다른 에이전트들과

DAML로 표현된 정보를 교류하여 획득한다.

ITTALKS상에서는 상호작용하는 다음 네 가지 에이전트들을 통하여 서비스가 제공된다.

1. ITTALKS 에이전트

사람과는 웹 기반 인터페이스로 반응하고, 다른 에이전트와는 ACL로 대화한다. IT 관련 세미나 정보를 제공한다.

2. 사용자 에이전트

사용자를 대신하는 비서의 역할을 수행하며, 사용자가 제시한 높은 수준의 목표(예: 참석 가능하고 관심사에 부합하는 세미나를 찾음)를 달성하기 위하여, 다른 에이전트들에 질의를 수행하고 정보를 선별한다. DAML을 이해하고, XSB를 이용한 지능적인 추론을 하며, ACL로 대화한다.

3. 분류 에이전트

신규 세미나 자료가 입력된 경우, ACM에서 제시한 주제 분류 체계에 따라 주제를 분류한다.

4. MapQuest 에이전트

세미나의 장소와 사용자의 현재 위치를 비교하는 기능을 수행한다.

세미나 관련 지식을 표현하기 위해서는 사용자의 관심사항, 세미나의 주제, 스케줄, 공간 및 시간, 에이전트의 신뢰도에 대한 온톨로지가 필요하다. 이 온톨로지들을 바탕으로 명시적으로 기술되지 않은 사실을 추가적으로 추론할 수 있다.

세미나 자료의 입력은 크게 세 가지 방식을 사용한다. 첫째, 웹 페이지에 정해진 필드들을 채워 문서를 작성한다. 둘째, DAML로 작성된 기 문서를 해석하여 내용을 받아들인다. 마지막으로 HTML로 작성된 기존 웹 텍스트에서 자동으로 세미나 정보를 추출

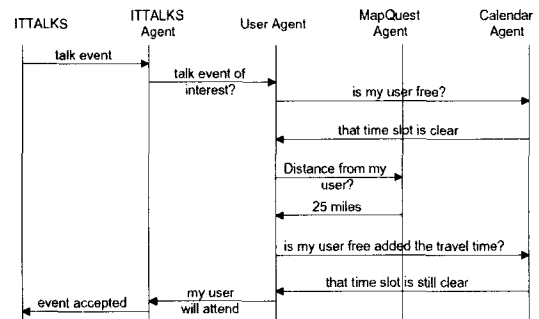


그림 3 ITTALKS에서 세미나에 참석 가능한 경우의 에이전트간 반응[2]

하여 입력한다. 일단 DAML로 세미나 내용이 작성되면, 다음 단계로 지식의 표현, 교환 및 전달에 폭넓게 사용할 수 있다.

3.5 그리드 컴퓨팅: 시맨틱 웹 기술의 테스트 베드[9]

입자 물리 실험과 같은 대규모와 장기간의 과학 실험에는 세계 여러 곳에 흩어져 있는 연구기관들이 공동으로 참여한다. 따라서 분산되고 이질적인 정보 시스템과 실험 장비들을 상호 연동하여 연구를 수행하여야만 한다. 최근에 부각되는 예로서 유전정보 데이터베이스의 활용을 들 수 있다. 생물정보학의 발전에 따라 이질적인 대규모 유전정보 데이터베이스를 연동하고, 수십 수백 대의 슈퍼컴퓨터를 동원한 대규모 계산을 수행할 필요성이 커지고 있다. 이 목적에 그리드 컴퓨팅 기술이 사용되고 있다. 이 기술을 이용하여, 인터넷을 통하여 계산 자원과 데이터를 마치 배전망처럼 필요에 따라 끊임없이 동적으로 할당받아 사용한다.

그리드 컴퓨팅은 시맨틱 웹 기술을 적용하기에 용이한 테스트베드이며, 거꾸로 그리드 컴퓨팅에 사용한 기술을 시맨틱 웹에 적용할 수 있다. 그리드 컴퓨팅의 실현을 위해서는 여러 기관에 분산되고, 시간에 따라 가용한 자원의 변화가 심한 환경에서 안정된 서비스를 제공하는 것이 필수적이다. 시맨틱 웹에서는 에이전트들이 서로 지식을 공유하며, 대규모의 에이전트와 사람이 참여하여 상호작용하며, 이질적인 시스템들이 상호 연동되어 작동할 것을 요구한다. 이 두 기술 모두 참여하는 시스템 자원의 자동 발견 및 통합이 필요하며, 이를 위하여 메타데이터의 처리가 필수적이다. 그리드 컴퓨팅을 지원하기 위해서는 다음 요구 사항을 처리하여야 한다.

첫째, 복잡한 계산을 처리할 필요가 발생한 경우, 그 시점에서 가용한 자원(문서, 도구, 인력)과 서비스(데이터베이스, 센서, 서버 등)를 동적으로 연합하여 구성할 수 있어야 한다.

둘째, 수백, 수천 개 이상의 서비스를 공통의 목적을 위하여 중앙의 통제없이 ad-hoc하게 구성하여 조절할 수 있어야 한다.

셋째, 몇 달 이상의 계산이 필요한 경우가 많기 때문에, 장시간에 걸쳐서 연합을 유지해야 한다.

이러한 요구조건을 만족하는 그리드 컴퓨팅에 적

용 가능한 시맨틱 웹 기술은 다음과 같다.

1. 추론을 이용한 필요 자원의 자동 발견 및 동적 구성

계산에 필요한 자원 및 서비스를 자동으로 검색하여 선택한 후, 호출하고, 실행 결과의 모니터링이 필요하다. 특성을 기록한 자원 목록으로부터 필요한 자원을 정밀하게 검색하기 위하여 추론 규칙을 적용하여야 할 것이다. 또한 급속히 서비스 환경이 변화하고, 자주 재구성되는 상황에서 불안정한 서비스는 다른 서비스로 대체할 수 있어야 한다. 보다 유연한 추론을 가능하게 하기 위하여, 자원 특성의 기술을 DAML+OIL로 하는 것이 필요하다.

2. 온톨로지를 이용한 이질적인 정보의 교환 및 통합

생물정보학의 유전 정보는 다양한 데이터 형태와 이질적인 데이터베이스에 저장되어 있다. 또한 저장된 유전정보 서술의 정밀도와 추상화 정도가 시스템마다 상이하다. 이러한 유전정보의 메타데이터는 온톨로지를 통하여 기술하고, 다양한 알고리즘과 함수들의 파라미터들을 RDF 등으로 기술하여 통합할 필요가 있다.

그리드 컴퓨팅과 시맨틱 웹기술 모두 궁극적으로 대규모로 사람과 에이전트가 참여하여, 서로 데이터 및 지식을 공유하고 처리하는 보편적인 환경을 개발하는 것을 목표로 한다. 그리드 컴퓨팅은 그 기술을 응용할 실제 현장의 문제를 갖고 있다. 시맨틱 웹 기술이 이 목적으로 사용될 수 있다. 즉 지식과 자원을 표준화하고, 이질적인 시스템을 결합하는데 공헌할 것이다. 반대로 시맨틱 웹 기술은 그리드 컴퓨팅에서 사용하는 검색, 변환, 대규모 자원 관리 도구들을 사용하여 발전할 것이다.

4. 시맨틱 웹의 이슈들

시맨틱 웹이 원래의 목적대로 기계가 의미를 이해하여 처리함으로써 사용자에게 편의를 제공하기 위해서는 다음 기술적, 사회적 문제를 해결하여야 한다.

첫째, 온톨로지의 생성, 통합, 수정 및 관리를 위한 도구가 필요하다. 각 기관과 사회마다 별도로 작성된 온톨로지는 효율적인 정보 공유를 위하여 통합할 필요가 있다. 서로 다른 어휘로 표현되는 온톨로지 개념과 이들의 관계를 변환하는 테이블이 필요할 것이다. 또한 온톨로지는 시간에 따라 변화한다. 예를 들

어 상품 분류의 온톨로지인 UNSPSC은 7개월마다 20% 가량이 변화한다고 한다[8]. 따라서 지속적으로 변화하는 온톨로지의 수정을 용이하게 하고, 버전을 관리하는 도구가 필요하다.

둘째, 사용자가 온톨로지에 기반한 문서를 쉽게 작성하도록 보조하는 도구의 개발이 필요하다. 자연어로 기술된 문서에 비하여 온톨로지에 기반한 문서는 사용자의 노력이 많이 든다. 문서를 작성하고, 특정 개념마다 온톨로지를 검색하고, 문서를 해당 온톨로지와 연결하는 것은 시간이 많이 걸리며, 오류를 일으키기 쉽기 때문이다. 따라서 편리한 온톨로지 문서 작성 도구의 개발이 필요하다. 또한 온톨로지 입력의 수고를 줄이기 위하여, 기존 자연어 문서의 내용을 기계로 분석하여 온톨로지에 연결하는 도구도 필요하다.

셋째, 온톨로지서 암묵적으로 내재된 새로운 지식을 실시간으로 추론하고, 유추한 지식을 관리하는 도구가 필요하다. 지식 추론은 사용되는 규칙의 수와 지식의 양에 지수함수로 비례하여 계산시간을 요구한다. 따라서 사용자에게 실시간으로 반응하는 웹 검색을 지원하기 위하여, 복잡한 개념을 실시간으로 추론하는 엔진이 필요하다. 또한 이전의 제한된 지식으로 추론한 지식은 새로운 사실이 밝혀지거나, 과거의 지식에 오류가 있다고 판명되면, 다시 추론이 이루어져야 한다. 어느 지식이 어느 지식에서 추론된 것인지 파악하고, 새로운 사실이 입력된 경우 추론을 전파하여 변경을 반영하는 truth maintenance 시스템이 필요하다.

마지막으로 사회적 이슈로, 시맨틱 웹의 확산을 장려하기 위하여 시맨틱 웹의 성공사례를 제시하는 것이 필요하다. 시맨틱 웹은 간단한 에디터로 작성이 가능하였던 기존 웹에 비하여, 작성이 번거롭고, 비용이 많이 든다. 따라서 기존 웹을 시맨틱 웹으로 바꾸는 비용에 비하여, 정보의 공급자가 변환후의 이익이 크다는 것을 납득할 수 있어야 한다. 한편으로 정보의 사용자가 시맨틱 웹 기술을 활용하여 얻는 이익은 시맨틱 웹으로 작성된 정보의 비중에 비례하여 증가한다. 따라서 정보의 공급자와 사용자 모두에게 시맨틱 웹 활용을 장려할 수 있는 사례(킬러 애플리케이션)의 소개가 필요하다.

5. 결론

시맨틱 웹은 기계가 정보의 의미를 이해하고 처리

할 수 있도록 기존 웹을 확장하는 것을 목표로 한다. 사람에게 기록된 정보를 보여주는 데 중점을 둔 기존의 웹 기술로는 기하급수적으로 규모가 팽창하며, 동적으로 정보가 생성되고, 사람과 기계가 함께 참여하여 상호작용하는 진화된 웹 환경을 지원하는데 한계가 있다. 시맨틱 웹에서는 문서, 정보, 개념 등을 유일한 이름(URI)로 지칭하고, 명시적으로 온톨로지를 통하여 지식을 공유한다. 기계가 정보의 내용을 잘 이해함으로써, 정밀한 정보의 선별, 분류 및 검색이 가능하고, 기계와 기계, 기계와 사람간 지식 공유 및 교환을 지원한다.

본 논문에서는 지식 관리 시스템, 전자상거래를 위한 상품의 체계적 분류, 음악 메타데이터, IT 관련 세미나, 그리드 컴퓨팅 등의 도메인에 시맨틱 웹을 적용한 사례를 살펴보았다. 현재 시맨틱 웹은 제한된 도메인을 갖는 분야에 우선 도입되어 가능성을 테스트하는 상황이라고 볼 수 있다. 앞으로 시맨틱 웹이 활발히 적용될 분야로 정보와 지식의 표현, 교환 및 검색이 중요한 분야, 웹 서비스와 같이 서비스 내용 검색 및 소프트웨어간 파라미터의 표준화가 중요한 분야와 에이전트간 지능적인 정보 교환이 중요한 분야를 들 수 있다.

시맨틱 웹은 앞으로 다양한 분야에서 더욱 활발히 활용될 것이다. 현 인터넷을 단순한 정보 흐름의 통신망에서 지식을 생산할 수 있는 지능적 그물로 진화시키는 작업의 시금석이 바로 시맨틱 웹이다.

참고문헌

- [1] 김홍기, 김학래, 이강찬, 정지훈, 이재호 외, "월드와이드웹에서 시맨틱 웹으로", 마이크로소프트웨어 시맨틱 웹 특집, pp. 242-301, 2002년 4월
- [2] R. Cost, T. Finin, A. Joshi, Y. Peng, C. Nicholas, I. Soboroff, H. Chen, L. Kagal, F. Perich, Y. Zou, S. Tolia, "TTTALKS: A Case Study in the Semantic Web and DAML," Int. Semantic Web Working Symposium, Stanford University, USA, 2001.
- [3] S. Bergamaschi, F. Guerra, and M. Vincini, "A Data Integration Framework for e-Commerce Product classification," The First International Semantic Web Conference, Sardinia, Italy, June, 2002.

[4] T. Berners-Lee, J. Handler, and O. Lassila, "The Semantic Web," Scientific American, May, 2001.

[5] T. Berners-Lee, "The Semantic Web," <http://www.w3.org/2002/Talks/04-sweb/>, Academic discussion talk, Japan Prize 2002.

[6] DARPA, DARPA agent markup language website, <http://www.daml.org>

[7] Y. Ding, M. Korotkiy, B. Omelayenko, et. al., "GoldenBullet: Automated Classification of Product Data in E-commerce," Business Information Systems Conference (BIS 2002), April, 2002.

[8] D. Fensel, C. Bussler, Y. Ding, V. Kartseva, M. Klein, M. Korotkiy, B. Omelayenko, and R. Siebes, "Semantic Web Application Areas," the 7th International Workshop on Applications of Natural Languages to Information system, Stockholm, Sweden, June, 2002.

[9] C. Goble, D. Roure, "The Grid: An Application of Semantic Web," ACM SIGMODE Vol. 31, No. 4, Dec. 2002.

[10] J. Hendler, T. Berners-Lee and E. Miller, "Integrating Applications on the Semantic Web," Journal of the Institute of Electrical Engineers of Japan, Vol. 122(10), pp. 676-680, Oct., 2002.

[11] MusicBrainz., MusicBrainz, <http://www.musicbrainz.org>, <http://logicerror.com/musicbrainzArticle>

[12] On-To-Knowledge project consortium., On-To-Knowledge project website, <http://www.ontoknowledge.org>

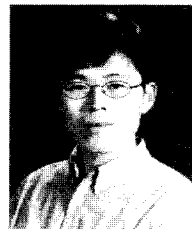
[13] Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer, D. Wenke, "OntoEdit: Collaborative Ontology Development for the Semantic Web," The First International Semantic Web Conference, Sardinia, Italy, June, 2002.

[14] UNSPSC., UNSPSC website, <http://eccma.org/unspsc>

[15] W3C. RDF website. <http://www.w3c.org/RDF>

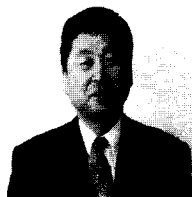
[16] W3C. RDFS website. <http://www.w3c.org/TR/rdf-schema>

조 성 정



1996 KAIST 전산학과 학사
 1998 KAIST 전산학과 석사
 2003 KAIST 전산학과 박사
 2002 7~11 미국 IBM Watson 연구소 연구 학생
 2003 삼성종합기술원 유비쿼투스 컴퓨팅 그룹 입사 예정
 관심분야 : 인공지능, 패턴인식, 기계학습, 문자인식
 E-mail : sjcho@ai.kaist.ac.kr

김 진 형



1971 서울대학교 공대 학사
 1979 UCLA 시스템 엔지니어링 석사
 1983 UCLA 전산학과 박사
 1985~현재 KAIST 전산학과 교수
 1995~1999 출연(연) 연구개발정보센터 소장
 현재 공학한림원 회원, 과학기술한림원 회원, IAPR Fellow
 관심분야 : 인공지능, 패턴인식, 베이지안 네트워크, 인공지능경망, 은닉마르코프 모델
 E-mail : jkim@ai.kaist.ac.kr
